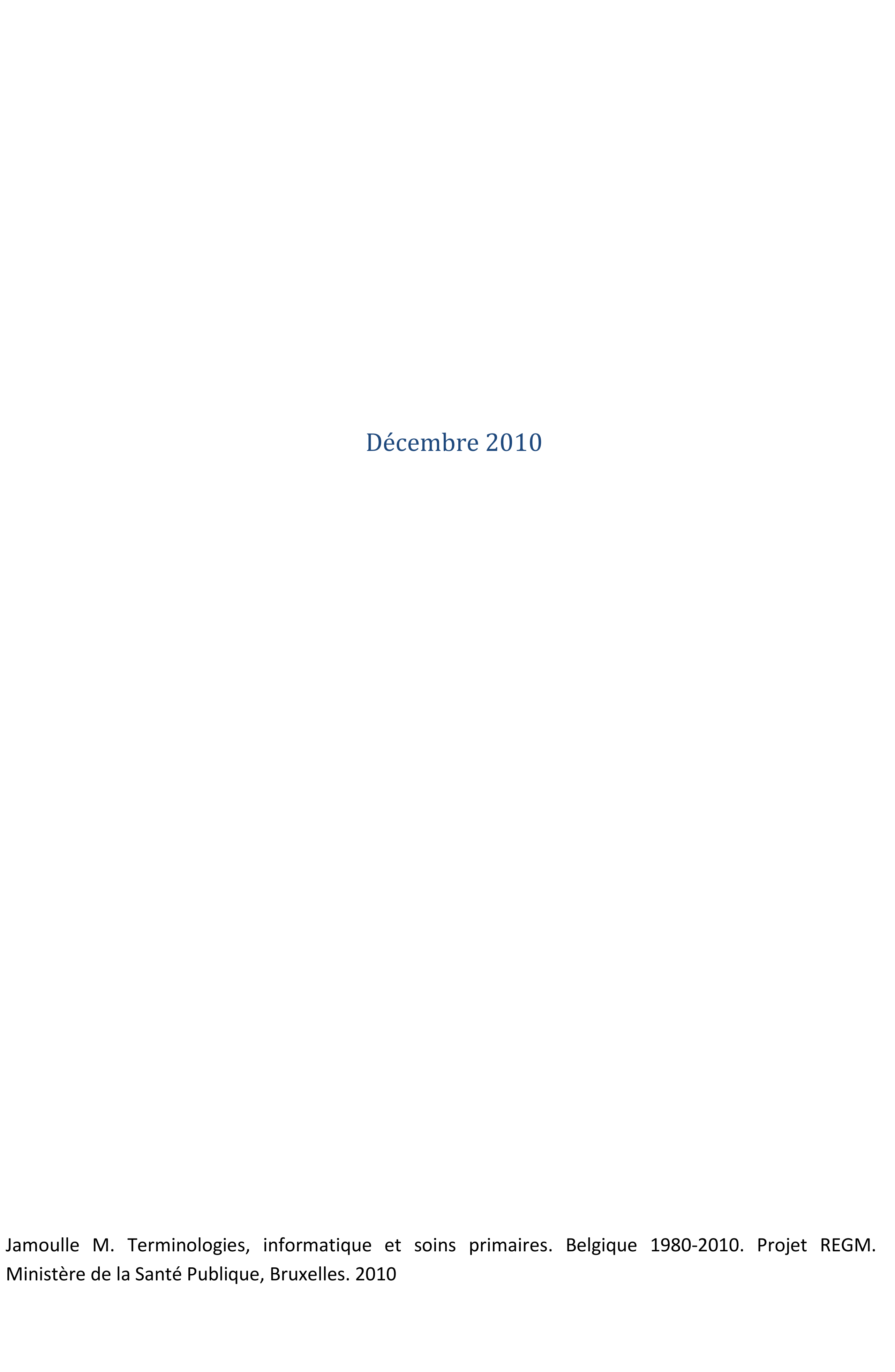
Terminologies, informatique et soins primaires.

Belgique 1980-2010

Projet REGM. Ministère de la Santé Publique, Bruxelles. 2010



Décembre 2010

Jamoulle M. Terminologies, informatique et soins primaires. Belgique 1980-2010. Projet REGM. Ministère de la Santé Publique, Bruxelles. 2010

Sommaire

Abstract

La mise en place de systèmes d'information de santé, cohérents et évolutifs, nécessite l'adoption de terminologies de référence dans le domaine de la santé. Nous examinons ici les différentes approches terminologiques disponibles dans les logiciels belges, l'effort d'une terminologie centralisée connue sous le nom de 3BT, la tendance internationale autour d'un produit anglo-saxon dénommé SNOMED-CT et nous indiquons la voie qui nous parait la plus adéquate pour le développement d'une informatique médicale plus fluide dans les années à venir, condition indispensable pour l'analyse ultérieure de données de morbidité.

1. Dictionnaires et terminologies en usage dans les logiciels belges de médecine générale

1.1.Standardiser les mots

La mise en place de systèmes d'information de santé, cohérents et évolutifs, nécessite l'adoption de terminologies de référence dans le domaine de la santé[1]. Cela parait être une évidence dans la bouche d'un professeur de santé publique en 2010 mais il a fallu plusieurs dizaine d'années pour s'apercevoir que les efforts des concepteurs de logiciel s belges pour munir leur produit de terminologies se traduisent par une cacophonie de dictionnaires non standardisés rendant infaisable l'agrégation du savoir indispensable à l'analyse.

Standardiser les mots n'est pas une tâche de docteurs et pourtant ce sont des médecins qui depuis les années 80 ont, souvent avec un admirable acharnement, développé des dictionnaires internes à leur produit logiciels puis essayé de les faire correspondre aux différentes classifications existantes.

La technique utilisée a été systématiquement conditionnée par l'informatique disponible. Quasi tous ont commencé en DOS et Dbase et ont suivi en ACCESS puis SQL et donc la fabrication de liste de mots fréquents puis moins fréquents puis rares puis tres rares est devenue une occupation à plein temps pour certains au fur et à mesure que le dossier médical informatisé se raffinait.

L'état belge a même sponsorisé pendant dix ans une équipe de chercheur pour établir une liste de mots encore plus longue.

Pendant ce temps l'informatique allait son chemin et les ordinateurs devenaient de plus en plus puissants, permettant des traitements d'information inimaginable jusqu'alors. L'émergence du langage OWL en 2001 a sonné le rassemblement des terminologues, documentalistes et informaticiens et autres spécialistes du Knowledge management autour de ce que Wuster appelait la Théorie Générale de la Terminologie[2] il a fallu encore quelques temps pour que les professionnels de la médecine emboitent le pas de cette nouvelle approche et considèrent ces nouveaux outils que le monde industriel a déjà engrammé depuis plusieurs années.

Nous examinons ici les différentes approches terminologiques disponibles dans les logiciels belges, l'effort d'une terminologie centralisée connue sous le nom de 3BT, la tendance internationale autour d'un produit anglo-saxon dénommé SNOMED-CT et nous indiquons la voie qui nous parait la plus adéquate pour le développement d'une informatique médicale plus fluide dans les années à venir.

On peut grossièrement distinguer trois étapes dans l'élaboration des terminologies. Les terminologies historiques, construite au moyen des bases de données des années 80 fournissent des listes d'occurrences selon le terme cherché par l'utilisateurs. On peut les appeler terminologie de première génération (Ex 3BT).

---

1 Fieschi M. Un enjeux, les terminologies médicales . sur http://esante.gouv.fr/pointsdevue/marius-fieschi-un-enjeu-les-terminologies-medicales

2 Wuster E. : "La théorie générale de la terminologie -un domaine interdisciplinaire impliquant la linguistique, la logique, l'ontologie, l'informatique et les sciences des objets", dans DUPUIS (H.), éd., Essai de définition de la terminologie. Actes du colloque international de terminologie (Québec, Manoir du lac Delage, 5-8 octobre 1975), Québec, Régie de la langue française, p. 49-57 , 1976. accessible sur http://docpatient.net/onto/doc/wuster.pdf

L'apparition dans les années 2000 de nouvelles techniques informatiques a premis d'introduire dans le développement terminologique des outils puissants de relation entre des concepts tout en gardant la structure de liste telle quelle. On peut les appeler terminologie de deuxième génération qui mixent les pick list et les relations ontologiques (ex SNOMED-CT)

L'apparition des nouvelles techniques informatique tel que le langage OWL et le web sémantique permet le développement d'outil terminologique nouveaux, puissants et utilisable avec les technique de Natural Language Processing. On pourra les appeler terminologies de troisieme génbération.

1.2.Enquète Prorec pour SEMINOP

L'enquête menée auprès de douze producteurs d'applications médicales membres de Prorec.be par le groupe SEMINOP du Ministere de la Santé a montré la grande variation d'utilisation terminologique par les logiciels belges[3]. On voit que seulement 6 sur 1é utilise 3BT, un n'utilise que la version 1987 de ICPC et 5 utilisent des dictionnaires propriétraires. Le tableau ci-dessous montre cette variété.

Terminologies		Diag	RFE
	3BT	6	5
Direct	ICPC-1	1	1
	propriétaire	5	4
Aucun		1	3
	ICPC-1	0	4
	ICPC-2	10	5
Indirect	ICD9-CM	5	0
	ICD10	10	5
	SNOMED	0	0

TABLE 1.1 – Enquête menée dans le cadre de PROREC-BE sur l'exploitation de terminologies et de leur mode (directement ou indirectement par l'industrie en Belgique et présentée lors du SEMINOP du 24 juin).

1.3.Les dictionnaires propriétaires

Sauf à reproduire des copies d'écran, il est quasi impossible d'obtenir des sociétés de soft des détails sur leur dictionnaire qu'elles semblent considérer comme des secrets industriels. Elles ne répondent même pas aux mails envoyés à ce sujet.

3 Piette Ph., Celis MT, De Clercq E, Devlies J, Jamoulle M, Roland M, Van Bruwaene B, Vandenbergh Ph, Vander Stichele R. Recherche préliminaire en rapport avec le développement d'un serveur belge de terminologie dans le secteur de la santé. SEMINOP. MIM. Bruxelles, octobre 2010

Medidoc

J'au pu obtenir du Dr Bartolomeeussen, maitre d'œuvre du projet Intego[4], une page extraite du serveur terminologique du software Medidoc à titre scientifique.

La liste ci-dessous est manifestement prise à la façon des terminologies de première génération en introduisant le mot clef NEO BRONCH qui retourne une série d'occurrences dont seulement quelques unes sont présentées ici. Ce genre de liste, analogue à celles présentées par le 3BT (cfr infra) sont généralement peu aisées à manipuler et ne sont gérables que grâce à des astuces d'écrans préparées par les informaticiens et qui introduisent souvent des biais de sélection. Il s'agit souvent des occurrences les plus fréquentes qui sont présentées par défaut ou des occurrences préférées d'un utilisateur.

MinVanMddc_Cod	Diagn_Mddc	ICPC_Cod
R16281A.	NEO BRONCHIOLO-ALVEOLAIR	R84
R16235..	NEO BRONCHUS BOVENKWAB	R84
R16232..	NEO BRONCHUS BOVENKWAB ADENOCARCINOOM	R84
R162312.	NEO BRONCHUS BOVENKWAB EPIDERMOID CARCIN	R84
R16233G.	NEO BRONCHUS BOVENKWAB GROOTCELLIG CAR	R84
R16233..	NEO BRONCHUS BOVENKWAB GROOTCELLIG CARCI	R84
R16234..	NEO BRONCHUS BOVENKWAB KLEINCELLIG CARCI	R84
R16239..	NEO BRONCHUS BOVENKWAB NNO	R84
R1626...	NEO BRONCHUS EN PARANEOPLASTISCH SYNDROO	R84
R1622...	NEO BRONCHUS HILAIR	R84
R16222..	NEO BRONCHUS HILAIR ADENOCARCINOOM	R84
R162244.	NEO BRONCHUS HILAIR KLEINCELLIG CARCINOO	R84
R16229..	NEO BRONCHUS HILAIR NNO	R84
RC340...	NEO BRONCHUS HOOFDTAK	R84
RC342...	NEO BRONCHUS MIDDENKWAB	R84
R16242..	NEO BRONCHUS MIDDENKWAB ADENOCARCINOOM	R84

Tableau 1 MEDIDOC Extrait de la terminologie Occurrences sur l'entrée NEO BRONCH (vue partielle)

Dans l'application , différents choix sont aussi possibles ainsi qu'un lien vers une banque de données qui détaille la pathologie et propose les codifications internationales :

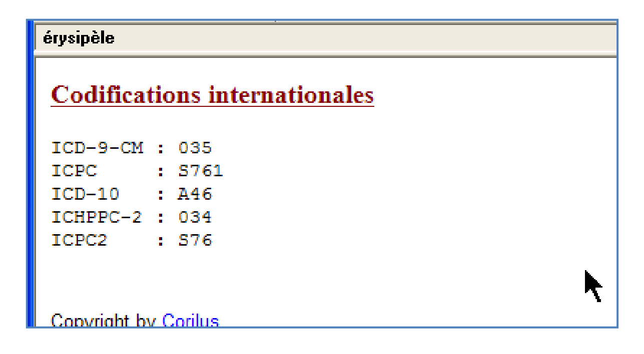


Tableau 2 Correspondance du terme Erysipele (Medidoc)

4 Bartholomeeusen S, Truyers C, Buntinx F. Intego, een registratienetwerk van huisartsen in Vlaanderen. Huisarts Nu 2005; 34: 11-14.

De Cantelis Medici

La banque de données DCM de 27.000 termes (De Cantelis Medici) fait partie intégrante de Medigest. En fait, DCM n'est pas "une" banque de données, il s'agit réellement de "multiples" banques de données : le sommaire offre la possibilité de gérer 62 banques de données en une seule : codes de 0 à 9, de A à Z et de a à z.

Environ 11.000 termes sont liés à ICD10 et 10.000 à ICPC2. La banque MED contient 21.000 termes dont 15.000 conditionnements sont liés aux codes CNK[5] avec une structure similaire à ATC.

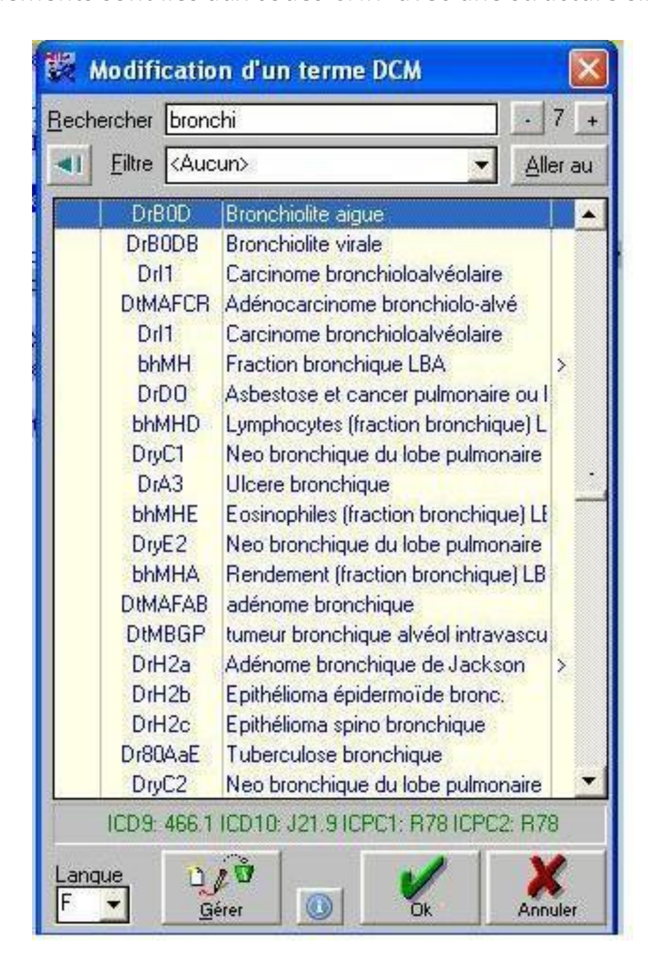


Figure 1 Lexique DCM, recherche des termes commençant par BRONCHI Le code choisi DrBOD Bronchiolite aigue correspondant à plusieurs codes (en vert au bas de la figure)

Au premier rang, les "codes à une position" du sommaire établissent une différence entre les plaintes ou interrogatoire, l'examen clinique, les diagnostics et les paramètres ainsi qu'entre les demandes d'examen, leurs résultats et divers domaines plus techniques ou administratifs. Au second rang, les "codes à deux positions" différencient, dans chacun de ces grands domaines, les diverses spécialités médicales. Ainsi, Diabète sucré (code DeB)" est un diagnostic (D) et endocrinologique (De) de surcroît. La terminologie DCM est transcodée à la CIM9, CIM10, CISP2 et DSM4. On voit par exemple dans la Figure 4 que le terme codé DCM 'DrBOD' est relié aux classifications ICD9, ICD10, ICPC1 et ICPC2.

---

[5] CNK ; A chaque nouveau conditionnement de spécialité pharmaceutique est attribué un code unique : le Code National ou Nationale Kode (CNK). Voir http://www.ipheb.be/annexe%20methodologique%20FR.pdf

Health One

Health one utilise une base de données terminologique propriétaire de 10.000 termes identifié par un contenu et un code (HO-Code). Chaque HO-Code est lié à un code ICPC2 et/ou ICD10 . Les codes diagnostics sont classifiés en une hiérarchie de maladies et en une liste de termes préférentiels.

1.4.Le LOCAS - Logiciel de Codage et d'Acquisition de Synonymes

Le LOCAS[6] est en réalité une base de données terminologique de première génération qui répertorie les termes les plus usités des médecins généralistes travaillant en Maisons Médicales entre 1990 et 2000. Ces termes sont transcodés à ICPC-2. Le LOCAS a fait l'objet d'une publication[7] . L'ouvrage est la transcription sur papier de la terminologie développée sous l'acronyme LOCAS (Logiciel de Codage et d'acquisition de synonymes).

Z01.0011	PROBLEME D'ARGENT
Z01.0021	PROBLEME DE PAUVRETE
Z01.0031	PROBLEME DE SURENDETTEMENT
Z01.0041	PROBLEME LIE A LA GESTION DE L'ARGENT
Z01.0051	PROBLEME ECONOMIQUE
Z02.0011	PROBLEME D'EAU
Z02.0021	PROBLEME DE NOURRITURE
Z03.0011	PROBLEME DE CHAUFFAGE
Z03.0021	PROBLEME D'HABITAT
Z03.0031	PROBLEME DE RELATION DE VOISINAGE
Z03.0041	VAGABONDAGE
Z03.0051	SANS DOMICILE FIXE
Z04.0011	DISCRIMINATION ETHNIQUE
Z04.0021	DISCRIMINATION LIEE A UNE GROSSESSE ILLEGITIME
Z04.0031	DISCRIMINATION
Z04.0041	DISCRIMINATION LIEE A LA RELIGION
Z04.0051	DISCRIMINATION LIEE AU SEXE
Z04.0061	PERSECUTION
Z04.0071	PROBLEME LIE A UNE CROYANCE
Z04.0081	PROBLEME CULTUREL
Z04.0091	PROBLEME LIE A L'IMMIGRATION
Z04.0101	PROBLEME LIE A UNE PRATIQUE SOCIALE
Z04.0111	PROBLEME LIE A LA RELIGION
Z04.0121	PROBLEME DE SOLITUDE
Z04.0131	PROBLEME DE SOCIETE
Z04.0141	PROBLEME LIE A UNE EXCLUSION SOCIALE
Z04.0151	PROBLEME D'HYGIENE
Z05.0011	PROBLEME OCCUPATIONNEL
Z05.0021	PROBLEME D'ORIENTATION PROFESSIONNELLE
Z05.0031	PROBLEME LIE AU TRAVAIL

Tableau 3 Extrait de LOCAS. Ouverture des codes CISP-2 Z01 à Z05

[6] Michel Roland, Marc Jamoulle & Bernard Dendeau. LOCAS 2, logiciel de codage et d'acquisition de synonyme basé sur la CISP version 2. Care Editions asbl, 2000

[7] Michel Roland, Marc Jamoulle & Bernard Dendeau, Approche Taxinomique en Médecine de Famille Care Editions asbl, 1996.

Le LOCAS comprend 4610 termes et syntagmes d'usage courant en médecine générale codés systématiquement sur la CISP-2. De nombreux médecins des Maisons Médicales belges dont le logiciel a été équipé de cette terminologie entre 1990 et 2000 environ ont contribué à son développement en renvoyant des commentaires, des corrections ou des demandes de synonymes. Le LOCAS, initié lors de la publication de la CISP-1 (1987) a été révisé après la publication de la CISP-2 et les codes sont donc ceux de la CISP-2 (1998). Le LOCAS est disponible en fichier de base de données.

Le mode d'emploi forme le volume 1 et est mis à la disposition des chercheurs et producteurs de logiciels de médecine générale. Bien qu'ils soient couverts par des droits d'auteur, ils sont utilisables librement à condition de ne pas les altérer et d'en citer la source. Le fichier téléchargeable sur le site http://trix.docpatient.net rubrique CISP comprend un fichier "lisez-moi" qui décrit ces modalités et les sources à citer en détail.

ICPC, Wonca, WICC

LOCAS2 est une base terminologique destinée à être utilisée en clinique de médecine générale. Les concepts sont ceux définis par le WICC, groupe de travail international de la Wonca, l'organisation mondiale de la médecine générale. Le Wonca International Classification Committee a catégorisé une série de concepts de prévalence forte en clinique de médecine générale, les a définis et les a munis de critères d'inclusion et d'exclusion. L'outil a été appelé International Classification of Primary Care (ICPC) et traduit en Français sous le nom de Classification Internationale des Soins Primaires (CISP) La version utilisée ici est la version 2 publiée en 2000 et à usage informatique. Ces concepts ont été associés à des codes en trois caractères, le premier alphanumérique pour l'un des 17 chapitres. Les deux suivants, numériques, pour les composants. Ces codes ont des caractéristiques mnémotechniques.

ICPC et ses différentes version et traductions appartiennent à la Wonca qui doit être avertie de l'utilisation des codes dans des systèmes informatique. La Belgique a acheté la licence d'utilisation.

Le LOCAS 2 est libre de droit et peut être utilisé à condition de citer les sources et les auteurs. Il ne peut être modifié. Toute utilisation en dehors de la Belgique doit faire l'objet d'un accord avec la Wonca.

1.5. Multilingual Glossary of technical and popular medical terms in nine European Languages

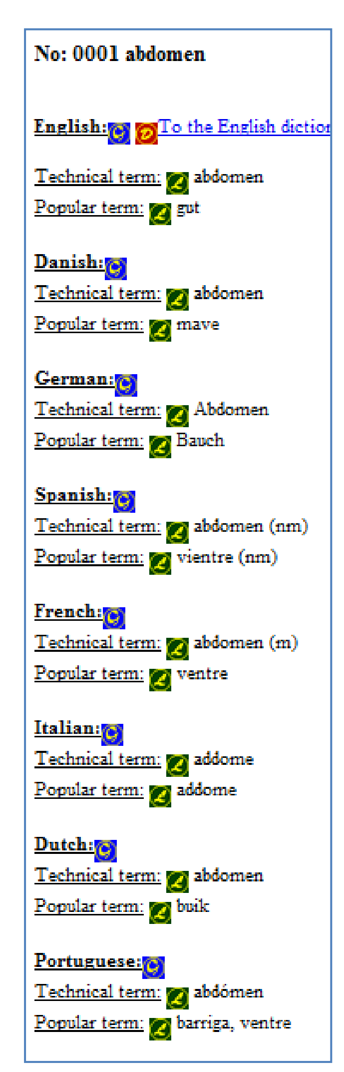


Figure 2 Multilingual lemma

Bien que cet outil ne figure pas directement dans les logiciels de médicine générale, il a une importance capitale à la fois pour l'exercice quotidien de la médecine et et pour des développements terminologiques futurs. En effet c'est sur base de ce dictionnaire que les notices boites des médicaments ont été adaptées à la lecture grand public et que nos patients peuvent comprendre le contenu des avertissements et précautions à prendre lors de l'usage de médicaments.

Ce projet a été soutenu à l'époque (1990-1995) par la Commission européenne et executé par le Heymans Institute of Pharmacology de l'Univbersiyté de Gand et le département de Linguistique appliquée de l'école Mercator[8].

Le systeme contient la forme électronique de 8 glossaires, soit 1830 termes médicaux populaires et techniques dans 8 des langues officielles de l'Union Européenne. (English, Dutch, French, German, Italian, Spanish, Portuguese and Danish).

Il y a quatre types de documents dans ce système :

1. La liste des termes médicaux techniques et populaire: ces listes contiennent un index de tous les termes médicaux, avec des références croisées entre les termes techniques et les termes populaires (une liste pour chaque langage).
2. Glossaire des termes médicaux techniques et populaire : ces glossaires contiennent des notes et des commentaires de traducteurs et utilisateurs pour chaque terme des 1830 dans le langage sélectionné.
3. Collection multilinguistique de lemmes[9] : Cette liste contient 1830 pages, une pour chacun des 1830 concepts dans ce système. On trouvera sur chaque page tous les termes techniques et populaires de tous les langages.
4. Dictionnaire : cette liste contient une description courte (en anglais) de chaque terme.

[8] Vander Stichele et coll. Multilingual glossary of technical and popular medical terms in nine European languages. Final Report. Heymans Institute of Pharmacology, University of Gent and Mercator College, Department of Applied Linguistics. Gent, Belgium. December, 1995 disponible sur http://users.ugent.be/~rvdstich/eugloss/information.html

[9] En linguistique, un lemme est une unité autonome constituante du lexique d'une langue (wikipedia)

List of technical and popular medical terms: French(p)

Back to welcome.

Back to language selection

Back to previous selection

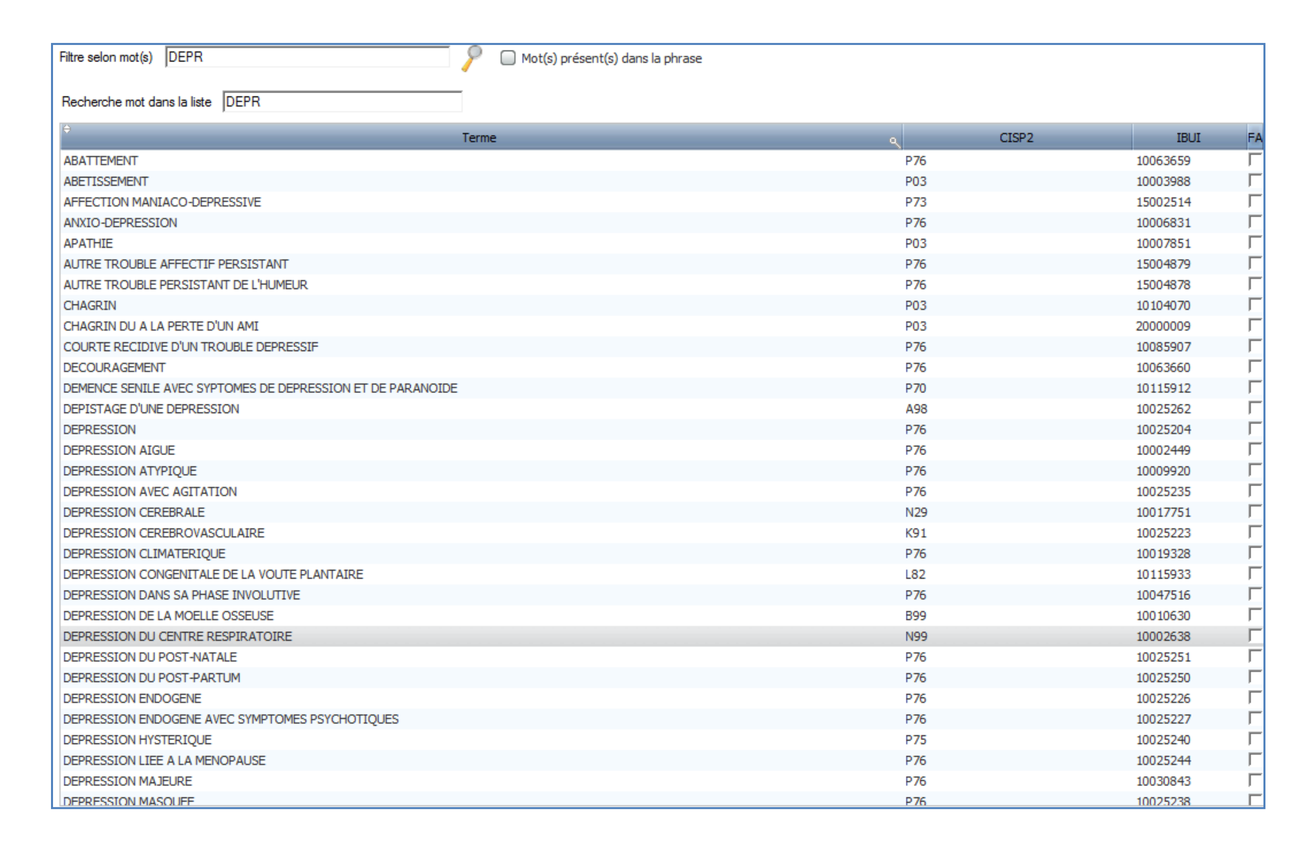
- **palliatif (s et am)** , *(traitement) qui agit sur les symptômes (sans agir sur la maladie)* (pop)
- **palpébral** , *des paupières* (pop)
- **palpitation (f)** , *contractions cardiaques anormales* (pop)
- **pancréatite (f)** , *inflammation du pancréas* (pop)
- **pancytopénie (f)** , *diminution simultanée des trois types de cellules sanguines* (pop)
- **panniculite (f)** , *inflammation du tissu adipeux sous-cutané* (pop)
- **papillaire** , *relatif à des papilles/en forme de-* (pop)
- **papillite (f)** , *inflammation de la papille optique/des papilles de la langue* (pop)
- **papillite/papillaire (oedème-)** , *inflammation de la papille optique (se traduisant par un oedème)* (pop)
- **papule (f)** , *bouton* (pop)
- **paracentèse (f)** , *ponction (pour évacuer un liquide anormal)* (pop)
- **paradoxal (sommeil-)** , *à activité cérébrale rapide (sommeil-)* (pop)
- *parallèle* (pop) , **parallèle**
- **parallèle** , *parallèle* (pop)
- *paralysie de l'accommodation de l'oeil* (pop) , **cycloplégie (f)**
- *paralysie légère ou incomplète* (pop) , **parésie (f)**
- **paralysie (f)** , *paralysie* (pop)
- *paralysie* (pop) , **paralysie (f)**
- **paramètre (m)** , *paramètre* (pop)

Figure 3 Multilingual Glossary. Extrait en Français)

1.6. 3BT , limites de la terminologie de première génération

1.6.1. 3BT développé sur base de l'expérience du LOCAS

Sur base de l'expérience du LOCAS, une collaboration entre l'Université Libre de Bruxelles (Dep de med gen Prof M. Roland) et L'université de Gand (Dep de med gen Prof De Maeseneer) soutenu annuellement par le Service Public de santé Fédéral a permis après plusieurs années de constituer une immense base terminologique de plus de 50.000 entrées , bilingue, codée sur ICPC et ICD10. Le produit a été présenté comme terminé en 2009 après de grosses difficultés méthodologiques et relationnelles entre les équipes francophones et néerlandophones. Le produit dénommé 3BT (pour Belgian bilingual biclassified terminology) représente un travail considérable et est un mix des termes retrouvés dans ICD10 et dans ICPC avec de nombreux synonymes, traduction du néerlandais en français et concepts composés.



Filtre selon mot(s) DEPR — Mot(s) présent(s) dans la phrase

Recherche mot dans la liste DEPR

Terme	CISP2	IBUI	FA
ABATTEMENT	P76	10063659	
ABETISSEMENT	P03	10003988	
AFFECTION MANIACO-DEPRESSIVE	P73	15002514	
ANXIO-DEPRESSION	P76	10006831	
APATHIE	P03	10007851	
AUTRE TROUBLE AFFECTIF PERSISTANT	P76	15004879	
AUTRE TROUBLE PERSISTANT DE L'HUMEUR	P76	15004878	
CHAGRIN	P03	10104070	
CHAGRIN DU A LA PERTE D'UN AMI	P03	20000009	
COURTE RECIDIVE D'UN TROUBLE DEPRESSIF	P76	10085907	
DECOURAGEMENT	P76	10063660	
DEMENCE SENILE AVEC SYPTOMES DE DEPRESSION ET DE PARANOIDE	P70	10115912	
DEPISTAGE D'UNE DEPRESSION	A98	10025262	
DEPRESSION	P76	10025204	
DEPRESSION AIGUE	P76	10002449	
DEPRESSION ATYPIQUE	P76	10009920	
DEPRESSION AVEC AGITATION	P76	10025235	
DEPRESSION CEREBRALE	N29	10017751	
DEPRESSION CEREBROVASCULAIRE	K91	10025223	
DEPRESSION CLIMATERIQUE	P76	10019328	
DEPRESSION CONGENITALE DE LA VOUTE PLANTAIRE	L82	10115933	
DEPRESSION DANS SA PHASE INVOLUTIVE	P76	10047516	
DEPRESSION DE LA MOELLE OSSEUSE	B99	10010630	
DEPRESSION DU CENTRE RESPIRATOIRE	N99	10002638	
DEPRESSION DU POST-NATALE	P76	10025251	
DEPRESSION DU POST-PARTUM	P76	10025250	
DEPRESSION ENDOGENE	P76	10025226	
DEPRESSION ENDOGENE AVEC SYMPTOMES PSYCHOTIQUES	P76	10025227	
DEPRESSION HYSTERIQUE	P75	10025240	
DEPRESSION LIEE A LA MENOPAUSE	P76	10025244	
DEPRESSION MAJEURE	P76	10030843	
DEPRESSION MASQUEE	P76	10025238	

Figure 4 3BT dans Epicure V14, recherche sur les digits 'depr. L'image montre une trentaine des 97 occurrences proposées.

3BT a été mis à disposition des concepteurs de logiciels informatisés et certains comme Epicure l'ont introduit comme base de référence terminologique. Sans développement d'interface utilisateur poussé et sans corrections ce produit terminologique fournit plus de bruits que de résultats. Interrogé selon les techniques de pick list (trois premiers digits ou premiers digits de deux termes qu'on pense associés) il fournit une liste impressionnante et absconse de terme en relation avec les trois lettres recherchées.

Ainsi dans Epicure (Version V14 de 2010) la requête sur Dépression au moyen des quatre digits depr fournit 97 (!) occurrences dont la pertinence terminologique et l'adéquation des codes est très souvent erronée. Un simple coup d'œil à la Figure 4 montre que Dépression cérébrale et Dépression cérébrovasculaire ne sont pas classées dans la même catégorie et cela probablement en raison d'une traduction erronée du néerlandais vers le français. On voit que toute statistique qui serait basée sur les codes issus de l'usage de cette terminologie serait sujette à caution.

1.6.2. Brève étude du contenu de 3BT (2009)

On reproduit ici une étude brève réalisée à compte personnel en 2009 par l'auteur de ces lignes.

The Belgian Biclassified Terminology is embedded in some EMRs in GP/FM in Belgium. I t is also being translated in Brazilian Portuguese by members of the Wonca International Committee The Belgian Evidence based medicine centre (cebam.be) and the recent project of registration in GP/FM are considering using it.

As an expert in classification and member of the WICC I have been asked at several moments to give comments on coding since 3BT has been developed but I didn't have the opportunity to have a good insight into the product.

There are the reason why I have begun to revise 3BT systematically at least in the most difficult components, the Symptoms and Complaints ones. Thus I have tried to analyze all the items proposed in 3BT code between 01 and 29 chapter by chapter.

After a quick insight in P, D and A chapter I have decided to continue systematically, beginning by the last one, the Z chapter. I have finished to revise 4 chapters; Z social Y and X : genitals and W : pregnancy.

The number of problems I have met is growing so much that I think it's necessary to share my analysis without finishing the 17 chapters.

Findings on 4 chapters:

Wrong code: 98 items of which 48 in the Z chapter (the code as been examined from the point of view ICPC and not from ICD one)

Basic term missing : terms like "manque d'eau" (lack of water) or "manque d'acces aux soins" (lack of health care accessibility) are missing

Term not understandable or inadequate or irrelevant: 92 of which 84 in the Z chapter with syntagmatic expression such as "*difficultés liées à la présence d'un parent handicapé dans la famille et provoquant une tension dans le foyer*" or moral affirmation such as "*simulateur pour des motivations évidentes*"

Objective facts instead symptoms or complaints: ICPC is not intended to code facts or objective findings like "*azoospermie"* (coded Y10) or *"status post-salpingo-ovariectomie"* coded X28

Disease instead symptoms or complaints: is already included in the wrong code rubric but the number of such error is so high that I have to mention it. Like "*sclérose prostatique*" coded Y06 or "exposition à une température extrême" coded Z29 (should be A88)

Process instead symptoms or complaints: this is particularly present in chapter W, partly due to inherent ICPC problem like « *examen de contrôle gynécologique dans le cadre d'un suivi de la contraception orale* » coded W11 (should be W31) but also in Y chapter as «test de fertilité chez l'homme » coded Y10 (should be Y38)

Bad translation Dutch/French ; is not frequent as in "*affection des organes génitaux*" for "*complicaties geslachtsorganen*"

Double coding needed : with 100 occurrences of which 94 in chapter W, this is clearly a point of interest to be discussed with the authors of 3BT. For instance the double syntagme "*hémorragie du post-partum due à un placenta incarcéré*" includes two concepts: one is related to symptoms and can be coded W17 post partum bleeding but the second is a medical diagnosis which has nothing to do here. The relationships quoted here are relevant of the structure of the information and not of the classification itself.

Considering this analysis of only 4 chapters which has been really time consuming, I suggest a complete process of revision of 3BT and I urge my colleagues from WICC who are in a translation process to participate to this revision before to use the product. The revision of 3BT has to be up to the Belgian Classification Committee with the help of all groups interested to have a good biclassified terminology for use in Primary Care. This kind of product, which has to be disseminated as an open document is a central issue in the development of EMR and particularly in open source EMR.

chapter	W	X	Y	Z	Total
change of code proposed	0	14	36	48	98
basic terms missing	2	8	7	11	28
term not understandable or inadequate or irrelevant	1	4	3	84	92
remarks and comments	10	14	3	10	37
objective facts instead symptoms or complaints	1	9	9		19
disease instead symptoms or complaints		3	10	25	38
process instead symptoms or complaints	36	0	17	0	53
bad translation Dutch/French	1	1	3	0	5
double coding needed	94	6	0	0	100
	145	59	88	178	470

Tableau 4 470 annotations on 4 chapters of 3BT (M. Jamoulle, 2009)

2. Les propositions récentes

2.1. Exemple d'une terminologie de deuxième génération ; SNOMED-CT

SNOMED-CT est un produit typique de la culture américaine. Produit au départ dans les années 70 par des anatomopathologistes , la terminologie de première génération (pick list) SNOMED (Systematized nomenclature of medicine) servait à identifier les échantillons histologiques. Elle est toujours utilisée dans ce sens par les laboratoires d'anapath. Il s'agit donc historiquement d'un classement de pièces sans vie. Au cours des trente dernières années le produit s'est développé pour répondre aux besoins essentiellement de l'industrie américaine des soins de santé dont on sait l'orientation hospitalière exclusive et la pauvreté de développement des soins primaires.

Dans les années 2000 SNOMED a évolué doublement, d'une part en s'adjoignant les codes Read, sensés représenter la clinique de la médecine générale anglaise et d'autre part en suivant le mouvement qui dans le domaine de la gestion de la connaissance et de l'intelligence artificielle a donné naissance aux ontologies ou du moins à la capacité pour les super ordinateurs de bureau d'enfin pouvoir traiter les concepts.

Dans une volonté d'hégémonie mondiale et on y voit en filligramme les besoins mondiaux de l'industrie de la santé, l'American College of Pathologist a lobyé de façon intensive, entrainant avec elle une série d'experts de différents pays et a réussi a lancer le label SNOMED-CT comme une marque au même titre qu'Yves St Laurent ou Adidas. Plusieurs pays ont accepté, parfois contre l'avis de leurs terminologistes de fonder une sorte d'institution internationale, aspirant au passage la

notion et le terme de standard (IHTSDO) à qui la CAP a revendu son produit pour une somme considérable tout en gardant la haute main sur l'organisation et le contrôle de qualité.

Ce qui frappe quand on assiste à un congrès SNOMED c'est l'ambiance marketing et business management. Quand on examine de prés le produit on voit qu'il garde les traces historiques de sa construction, quadrillé par 19 grandes parties qui sont censée recouvrir l'ensemble des besoins de santé. Il répond probablement aux besoins de l'industrie américaine de la santé mais malgré tous les efforts de diffusion, il n'est encore utilisé nulle part avec succès sauf en Argentine par une compagnie d'assurance privée qui possède ses hôpitaux.

Du point de vue des soins primaires, malgré une tentative de rapprochement avec la Wonca, l'organisation des infirmières ou l'OMS, on a de la peine à voir comment SNOMED pourra articuler le complexe de la médecine de premier recours qui croise les champs du somatique, du social du psychologique et de l'anthropologique.

Il faut souligner que les aspects mentaux de SNOMED sont un calque du DSM IV de l'Association Américaine de Psychiatrie dont les liens avec l'industrie pharmaceutique ne sont plus un secret pour personne. L'APA s'est acharnée à détruire le courant freudien aux Etats-Unis et à créer des nouvelles entités maladies correspondants aux besoins de cette même industrie qui payait ses travaux.

Cette critique acerbe n'enlève rien au fait que SNOMED est une terminologie hospitalo centrée remarquable et qu'on peut certainement tirer profit de cette somme de connaissance accumulée. Elle ne peut toutefois pas être un modèle terminologique pour les soins paires, y compris aux USA.

L'auteur de ces lignes a publié une étude détaillée sur SNOMED-CT. Cette étude est pertinente aux choix actuels et est reproduite en annexe de ce rapport pour le lecteur intéressé. Elle est aussi disponible sur Internet[10]

2.2. Vers une « Belgian end-user terminology »

Nous n'avons analysé ici que les aspects soins primaires des terminologies disponibles. Le récent rapport[11] du groupe SEMINOP[12] auquel l'auteur de ces lignes a participé étudie plus largement toutes les classifications et terminologies en usage en Belgique, quelque soit le niveau de soin. Le rapport SEMINOP conclut a une énorme disparité d'outil qui se superpose aisément ou font double emploi. Un long et conséquent travail d'harmonisation doit se mettre en place qui fera la part belle aux outils du web sémantique, préparant l'informatique médicale de demain.

2.3. Annexe; Some views about SNOMED-CT

[10] Jamoulle M. Some views about SNOMED-CT by a General Practitioner [Internet]. SEMINOP; Jul 2010. Available from: http://docpatient.net/onto/doc/SNOMED_CT_study_MJ_2010.pdf

[11] Seminop : Rapport final disponible sur http://www.health.belgium.be/eportal/Healthcare/Telematics/Studygroups/SEMINOP/index.htm?fodnlang=en

[12] Groupe SEMINOP voir https://portal.health.fgov.be/pls/portal/docs/1/10732006.PDF

Marc Jamoulle

Family doctor
Health data management specialist
WICC member
CAMG UCL marc@jamoulle.com

July 2010

Report to SEMINOP (Belgian Ministry of Health)

Some views about SNOMED-CT by a General Practitioner

Abstract

SNOMED- CT is is a concept-oriented and machine-readable medical terminology which has gained popularity this last ten years. It has been proposed as the reference terminology for use in electronic medical records and is supposed to cover the entire field needed to care and cure. This paper is examining its complexity and ability to cover the fields of general practice/family medicine, presented as the focus point of information in medicine.

Dr Marc Jamoulle is a free lance researcher. This work has been supported by a travel grant of the Centre Belge d'Information Pharmacothérapeutique (C.B.I.P.). Belgium http://www.cbip.be

Sommaire

1. Semantic interoperability and general practice

This paper is driven by the acquisition of knowledge in the field of semantic interoperability. This involves the ability of two or more computer systems to exchange information and have the meaning of that information accurately and automatically interpreted by the receiving system. In the field of interoperability, user's interface and consequently language is one of the main issue of the relationship with the computer.

Basic information about SNOMED CT is difficult to find, material available[13] lacks complete understandability by other than experts in formal terminology. Moreover SNOMED CT is a complex product, being the meeting point of medicine, computer sciences and computational linguistics. SNOMED-CT, a mix between US and UK terminologies, represents also a move from the terminological field to the ontological one and it is not so easy to understand the concept of ontology. One view of ontology is that it is a formal specification of a conceptualization for the purpose of enabling knowledge sharing and reuse[1]. This implies that relations between concepts are as important as the concept themselves.

Understanding the domain of terminologies, ontologies and their relationships, particularly of SNOMED- CT, represents a challenge for the general or family practitioner who is progressively facing the charge to gather in his EMR all the information generated by the health system for a particular patient. One can say that Information is to general practice / family medicine as technology is to medical specialties. Managing information is a new duty for the GP/FM, a complex domain as is General Practice itself, a crossroad between human and collectivity, between to do and to be[2], between patient interests and public expenses. If biomedical field is the first training step and experience, reality drives the practitioner through the depths of this symbolic animal called 'human being' (psychodynamics). This one lives in community and quantitative (epidemiology) and qualitative (anthropology) interactions are daily modifiers of practice. Consequently GP/FM is also the crossing point of Narrative Based (NBM)[3] and Evidence based Medicine (EBM) Let us assess whether SNOMED –CT's complexity is able to render the GP/FM one.

	To do	To be
individual	Bio math medicine	Mental health
community	Epidemiology	Anthropology

Figure 5 Four fields of General Practice / Family medicine
Adapted from M. Van Dormael[2]

[13] This paper has been written form material received at SNOMED CT® Training Class , April 27, 2010 during the IHTSDO conference in Copenhagen and from several Internet sources.

2. About SNOMED-CT & IHTSDO (nickname itsido)

SNOMED-CT means Systematized Nomenclature of Medicine - Clinical Terms. SNOMED is patented as a system and method for medical language extraction and encoding. SNOMED-CT is a standardized clinical cerminology, comprised of codes, terms and relationships, for use in recording and representing clinical information across the scope of health care which allows data to be captured (coded), retrieved, exchanged and understood. SNOMED-CT has been created by combining[4] SNOMED RT, the terminology produced by the College of American Pathologists (http://www.cap.org) and a computer-based nomenclature and classification known as Read Clinical Codes Version 3 (CTV-3)The Read Codes were invented and developed by Dr James Read in 1982, for use in General Practice. In 1990, the NHS bought the Read Codes from Dr Read and made them NHS standard. *Systematized Nomenclature of pathology* (SNOP), ancestor of SNOMED, has been published in 1964.

In the Medical subject heading of the National Library of Medicine, SNOMED has gained its place between MeSH and UMLS. It has been proposed – but not yet accepted - as the interface vocabulary for all the Electronic Health Records in the USA. Since 2003 there is an agreement between the U.S. Secretary of Health and Human Services, and the College of American Pathologists that will make SNOMED-CT available to U.S. users at no cost for the users, but a unknown cost to the US government through the National Library of Medicine's Unified Medical Language System® (UMLS).

In April 2007 the International Health Terminology Standards Development Organization , a non profit organization based in Denmark (http://www.ihtsdo.org), acquired the intellectual property rights of SNOMED Clinical Terms (SNOMED CT) and its antecedents from the College of American Pathologists for $7.8 million.

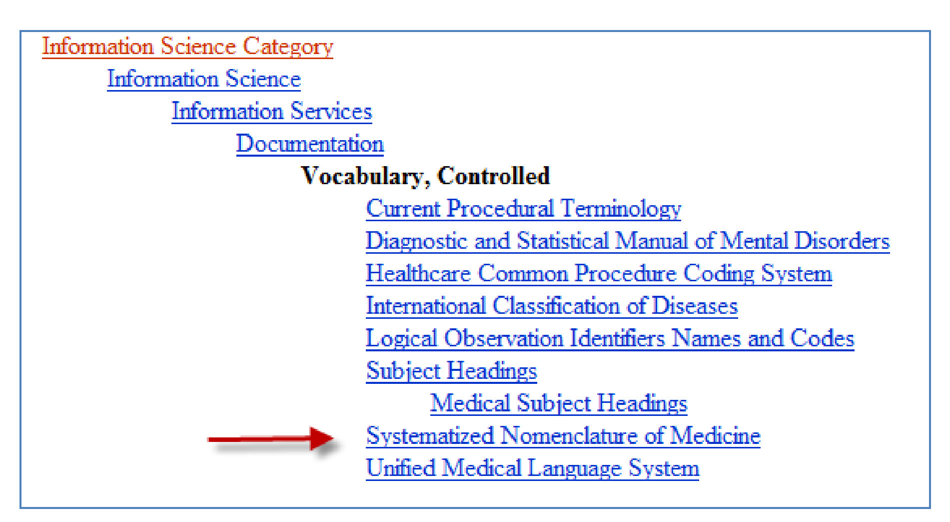


Figure 6 SNOMED finds its place in the controlled vocabularies of the NLM (http://www.ncbi.nlm.nih.gov/mesh/68018875)

2.1. International Health Terminology Standards Development Organization

By acquiring the SNOMED CT standard, the IHTSDO, which was established by a group of nine founding nations (Australia, Canada, Denmark, Lithuania, The Netherlands, New Zealand, Sweden, the United States and the United Kingdom), will help to ensure the continued maintenance and

evolution of SNOMED CT as well as its availability on an international scale. The IHTSDO assumed responsibility for the ongoing maintenance, development, quality assurance, and distribution of SNOMED CT. SNOMED Terminology Solutions, a specialized team of the College of American Pathologists has been commissioned to support IHTSDO operations under an initial 3-year contract and will separately continue to provide SNOMED-related products and services as a licensee of the terminology.

2.2. IHTSDO member countries (2010)

Spain	Australia	Lithuania
Sweden	Canada	New Zealand
The Netherlands	Cyprus	Singapore
United Kingdom	Denmark	Slovak Republic
United States	Estonia	Slovenia

3. SNOMED CT, a complex controlled medical vocabulary

3.1. What's a SNOMED concept

A unit of meaning

In SNOMED CT, a concept is a "Unit of meaning", described by a unique numeric code, a unique name (Fully Specified Name), a set of terms (descriptions), one "Preferred Term", and one or more Synonyms. SNOMED CT concepts are either "primitive" or "fully defined". A fully-defined concept can be differentiated from its parent and sibling concepts by virtue of its relationships with other concepts. Otherwise it is primitive.

Fully Specified Name

Unambigous name of concept, unique for active concepts which includes a trailing "hierarchy tag" in brackets. Ex;

– Hematoma (morphologic abnormality)
– Hematoma (disorder)

3.2. Descriptions versus definitions

A concept definition is the list of its relationships to other concepts[5]. There is no cursive definition of concepts available. While SNOMED-CT considers relationships as the unique way to define a concept, other authors of ontologies don't agree with this view, arguing that the ontologies have to include textual *definitions* for all terms and that terms should be defined so that their precise meaning within the context of a particular ontology is clear to a human reader.[6]

Indeed definitions by relationships could quickly become tautological as in the Figure 7. It's interesting to see the lack of stability of the relations used to describe interrelated concepts which are at the aim of the practice[7] The relation is_a normality findings could also be interpreted as part_of normality findings and one can wonder why normal is_anormality findings while abnormal is_a modifier. Indeed the question of relations in ontologies is still discussed

SNOMED-CT Term	# occurrences	Definition (by relations)
Borderline	188	is_a reference range interpretation value
Normal	1814	is_a normality findings
Abnormal	788	is_a modifier related to clinical specialty AND/or occupator
Pathological	92	is_a normality findings

Figure 7 Occurrences of usual terms in Clinical Clue Browser. (Cliniclue browser extract SNOMED CT International Edition (2010-01-31) http://www.cliniclue.com)

(WC: the above is of course nonsense)

3.3. The hierarchies or Roots of SNOMED-CT

They are 19 Upper level hierarchies or roots revealed by the Cliniclue browser. Each hierarchy (roots) has sub-hierarchies

Hierarchy | Subtype hierarchy

138875005 SNOMED CT Concept
- 123037004 body structure
- 404684003 clinical finding
- 308916002 environment or geographical location
- 272379006 event
- 106237007 linkage concept
- 363787002 observable entity
- 410607006 organism
- 373873005 pharmaceutical / biologic product
- 78621006 physical force
- 260787004 physical object
- 71388002 procedure
- 362981000 qualifier value
- 419891008 record artifact
- 243796009 situation with explicit context
- 48176007 social context
- 370115009 special concept
- 123038009 specimen
- 254291000 staging and scales
- 105590001 substance

Figure 8 The 19 hierarchies of SNOMED-CT concepts with their unique identifiers (Cliniclue browser extract SNOMED CT International Edition (2010-01-31) http://www.cliniclue.com)

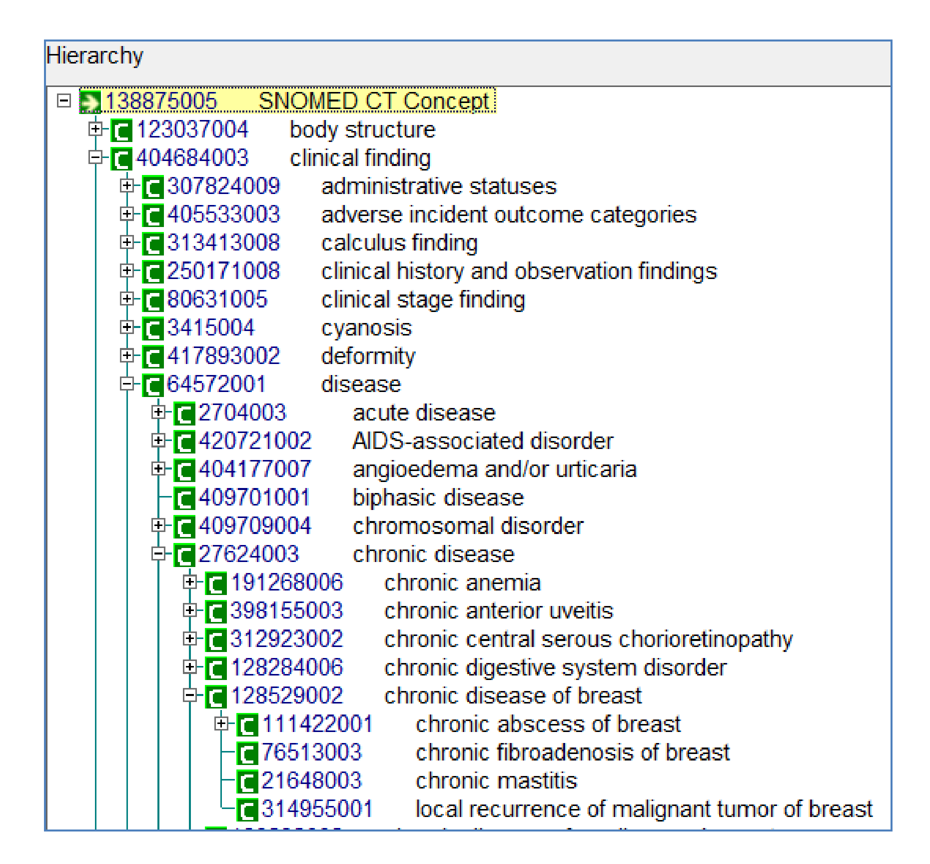


Figure 9 SNOMED-CT: Disease as sub-hierarchy of Clinical finding. The fig shows the opening of chronic disease of breast. (Cliniclue browser extract SNOMED CT International Edition (2010-01-31) http://www.cliniclue.com)

3.4. Relationships

Controlled vocabularies can be conceived as graph-theoretical structures consisting on the one hand of terms (which form the nodes of each corresponding graph) linked together by means of edges called relations[8].

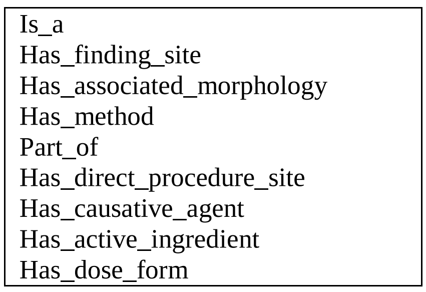


Figure 10 frequent relations in SNOMED CT

In SNOMED-CT, *Is_a* links concepts within a hierarchy while Attribute relationships allows links between concepts across hierarchies .

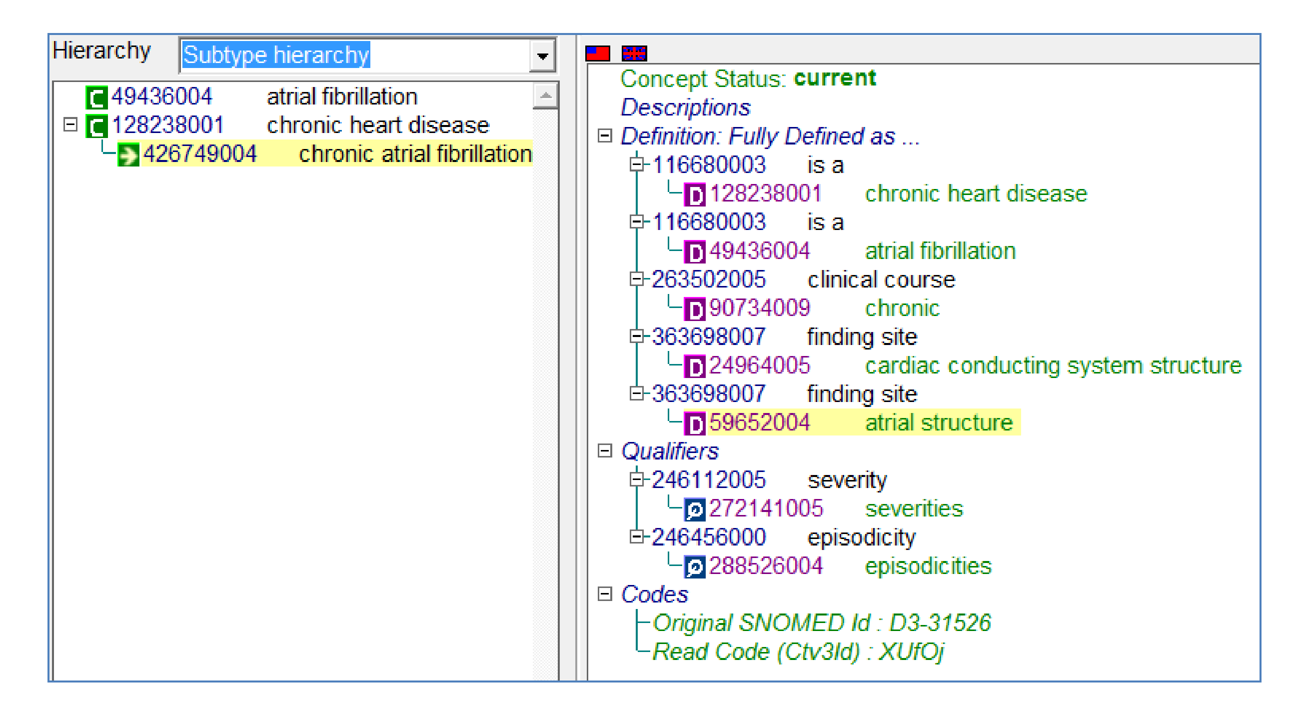


Figure 11 SNOMED-CT : left ; the concept Chronic atrial fibrillation is a sub hierarchy of clinical finding. Right ; relationships of the concept within a hierarchy (is_a) or across hierarchies including mapping to original codes. (Cliniclue browser extract SNOMED CT International Edition (2010-01-31) http://www.cliniclue.com)

3.5. The tables of SNOMED CT

SNOMED CT comprises 9 tables of which 3 core tables (concepts, descriptions and relationships) collectively defining a compositional description logic ontology of the medical domain, and a lexicon of associated preferred or synonymous descriptions. The most recent international release (January 2010) contains more than 308,000 active concepts, 791,000 active descriptions, organized into 19 hierarchies with more than 951,000 relationships

- 3 : Concepts, Descriptions, Relationships (Core tables)
- 2 : Subsets, Subset Members
- 3 : Cross Map Sets, Cross Maps, CrossMapTargets
- 1 : Component History

Figure 12 9 tables in SNOMED CT distributed in the form of tab delimited text files

Core tables characteristics:

Concepts	Relationships	Descriptions
Concept_id Concept Status Fully Specified Name CTV3_id SNOMED_id Is_Primitive	Relationship_id Concept id1 RelationshipType Concept id2 CharacteristicType Refinability Relationship Group	Description_id Concept_id Description Status Term Intial Capital Status Description Type Language Code
Gastric ulcer	Is_a	Gastrointestinal ulcer

Figure 13 Core tables characteristics

3.6. The Clinical Terms Identifier (SCTID), a 64 bit integer

Each '_id' is represented by a 64 bit integer (not always presented in full length), the SNOMED Clinical Terms Identifier (SCTID) like "101291657896189 00 9" (the last digit being a Verhoeff algorithm control code) used for primary key for most tables. The two digits before the last one are 10 for concept, 11 for description, 12 for a relationship but this form of presentation is rather uncommon. For example the Concept_id 22298006 means *myocardial infarction*

The concept-id could contain a Namespace Identifier, a code or that part of the 64 bit integer that identifies the organization responsible for creating and maintaining a standards-based extension or a standards-based derivative. Namespace Identifiers are 7-digit numbers that IHTSDO issues to those who create extensions to SNOMED CT,

The ConceptStatus is quoted by a number :

0 Current
1 Retire without specified reason
2 Duplicate
3 Outdated
4 Ambiguous
5 Erroneous
6 Limited/Classification
10 Moved Elsewhere
11 Pending Move

Example: the 3 following SCTID (Figure 14) are linked in such way that the Relationship_id identifies the chain of knowledge between two concepts

readability	Concept id1	RelationshipType	Concept id2
human	Cholecystectomy (procedure)	is_a (attribute)	Biliary tract excision (procedure)
machine	302619004	116680003	38102005
human	Gastric ulcer (clinical finding)	Is_a (attribute)	Gastrointestinal ulcer (clinical finding)
machine	397825006	116680003	40845000

Figure 14 SNOMED-CT ; Concepts and relations expressed in readable and coded form.

3.7. Pre and post coordination

SNOMED CT is compositional, allowing concept expressions to be pre-coordinated within the terminology or post-coordinated within the medical record.

Pre coordinated concepts

2474726011	pan retinal photocoagulation for diabetes
2533101018	attended diabetes structured education program
125663012	lipodystrophic diabetes with partial lipoatrophy
356077018	malnutrition-related diabetes mellitus - fibrocalculous
408810010	DTSQ - Diabetes treatment satisfaction questionnaire

Figure 15 Five of the 877 occurrences of the chain 'diabet' in Cliniclue browser. SNOMED CT International Edition (2010-01-31).These are typical example of pre coordinated concepts

Post coordination of concepts

SNOMED CT enables complex descriptions. For example, there might not be an explicit concept for a burn between the toes. But it could be described as

```
284196006|Burn of skin|:
 246112005|Severity|=
  24484000|severe,
   363698007|Finding Site|=
    113185004|Structure of skin between fourth and fifth toes|:
     272741003|Laterality|=
      7771000|left)
```

Figure 16 SOMED interpretation of burn between the toes. Example of post coordination.

Methods have been developed to retrieve pre- and post-coordinated concepts. The process for selectively retrieving pre- and post-coordinated SNOMED CT concepts relies on a syntactic transformation into a standard form[9].

Loosing redundancy

	ICPC	**ICD**	**SNOMED-CT**
Signifier	**B25**	**i64**	**397825006**
Signified	**Fear of AIDS**	**Stroke**	**Gastric ulcer**

Figure 17 Loose of redundancy in signifier / signified relationship

Machine readable implies that human beings cannot reasonably retain the code of the concept In the Figure 17 , the signifier is the code while the signified is the concept. ICPC has been designed to be used in Paper and Pencil format and incorporate mnemotechnic tricks. For instance B stands for Blood and in the whole classification the number -25 is reserved to Fear of. A GP who is familiar with the International Classification of Primary care (ICPC)[10] could quickly recognize the B25 code, in the context, as Fear of Blood disease of which the main one is Fear of AIDS. ICD10 code are not

mnemotechnic and a normal doctor will not retain that i64 is the code for Stroke while a dedicated coder could remember it. But the number 397825006 can only be ‘understood” by a computer. One remarks that going from left to right, one looses redundancy. While enhancing its precision ,he code looses internal information which allows human readers to understand it.

4. SNOMED Browsers

SNOMED CT is a complex ontology and sophisticated browsers are required to make it understandable and useful. Rogers & Bodenreider[11] have identified 23 SNOMED CT browsers that have been developed, and inspected 17. They enumerate and provide test criteria for a ‘master list’ of 143 browsing features supported by at least one inspected browser; future work will determine which of these features are implemented by individual browsers. Only 5 features were common to all 17 browsers; 89 were found in less than one third of browsers. They recommend that a core set of browsing features be defined and harmonized across browsers, particularly for text-to-concept .
Browsers permit searching on the SNOMED-CT terminology by displaying the hierarchy (descendents and ancestors) of a given concept and by displaying the relationships and terms (“detail”) associated with a particular concept search operations. Different implementations of software, offering SNOMED CT browsing capability are available. Do refer to http://www.connectingforhealth.nhs.uk

4.1. An example of study with Dataline software

The dataline SNOMED-CT browser allows visual representation (http://snomed.dataline.co.uk/) and is freely available on Internet. One has studied how a concept, in this case Epilepsy, is presented in serial written way with its concept_ids. (left column) and in a visual way (right column) (Figure 19)

The Epilepsy concept has the number 84757009 and pertains to ”(disorder)” hierarchy. The chain of character “epileps” is present 216 times in SNOMED-CT International Edition (2010-01-31) (Cliniclue browser) The Figure 19 shows some of them with their respective numerical identifier. For pedagogic purpose, one of the linked concept “Reflex epilepsy (disorder)” has been represented in the visual display of the left windows with its numerical identifier ”ID 79745005”. This helps to understand that the computer plays with numbers and shows readable items. The lines between concepts are actually relationships which have also their own numerical ID. The concepts are presented here with human readable words but for computer use chains of numerical ID are enough as shown in Figure 14Figure 14. Note that in the upper line of the left window one can read the SNOMED RT (CAP codes) and CTV3 (former Read code or clinical codes UK)

This kind of formal representation of concept forms the basis of ulterior transfers or analysis. XML is particularly fit to such transfer as in this example (Figure 18); the xml tag <description id> contains all the characteristics of the Epilepsy as a SNOMED-CT entry.

```
<description id="178739011" term="epilepsy" status="0" type="1" language="en"
initialCapital="0" conceptId="84757009" conceptStatus="0"></description>
```

Figure 18 XML tag for SNOMED-CT concept Epilepsy

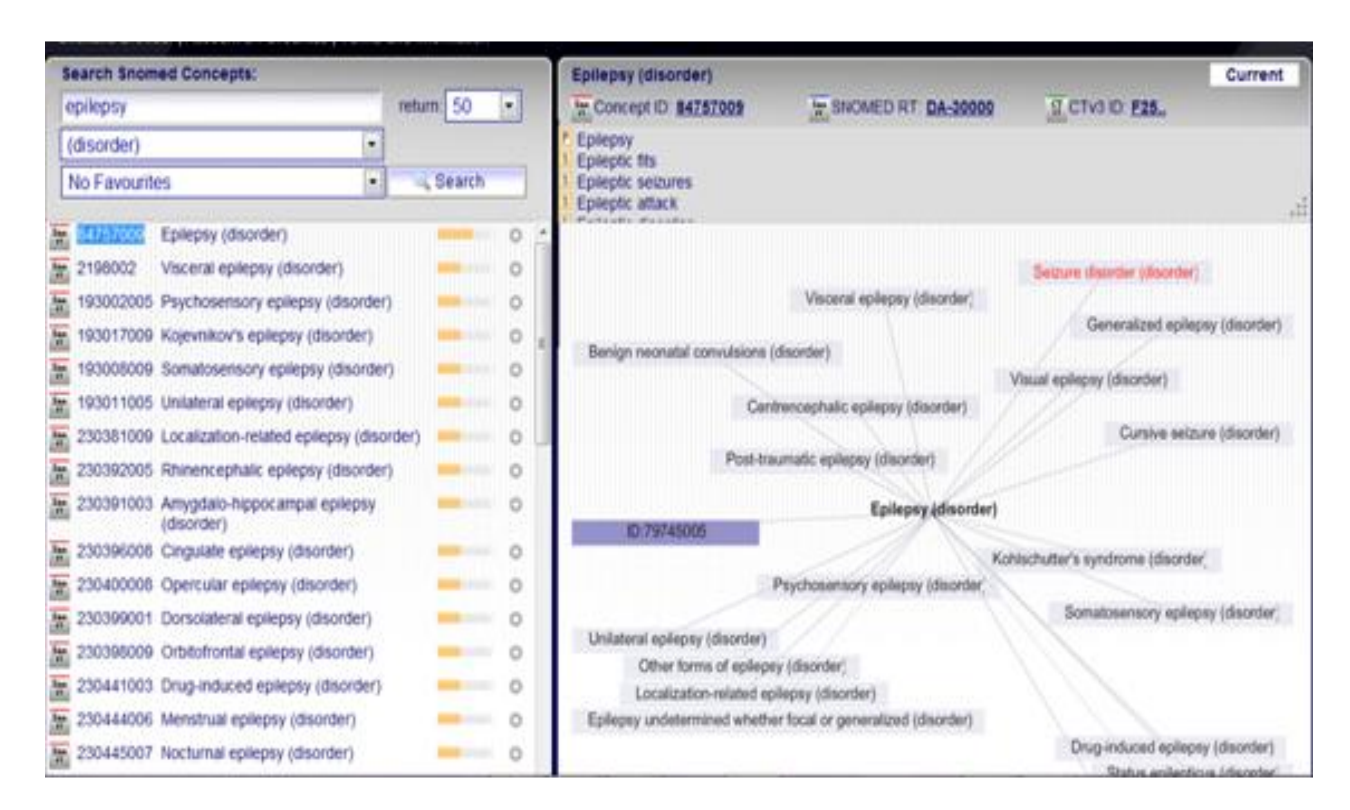


Figure 19 Snomed browser : the occurrence of the concept 'Epilepsy" with visual presentations of its relationships using Snoflake, the Snomed online CT browser from Dataline Software (http://snomed.dataline.co.uk/)

5. To what extent Is SNOMED-CT in use in practice?

It is unclear if SNOMED is really useful in clinical practice. A review of the available literature, performed in 2006 and published in 2008 examines 250 published studies on SNOMED. These studies look particularly relevant for proving the value of SNOMED in theory. Few studies are available on the usage of SNOMED in clinical practice[12].

An interesting study[13] , but yet outdated, compares the usefulness of 2.739 clinical codes of a dictionary, ICPC based, developed by Bob Bernstein (Canada) versus the 140.000 SNOMED 1993 entries. "The SNOMED coding specialists took more than 250 hours to encode 2,739 (ICPC based) codes (to SNOMED Xref) and the author took over 100 hours to revise them"

Practical use

Practical use seems not so easy. A university based hospital of 650 beds, serving 140.000 patients has implemented SNOMED-CT recently[14]. The authors describe three levels of implementation, which implies heavy and time consuming work for skilled professionals in IT and terminology.

- Entry Terminology: this is the user interface, the terms used to interact with users during data entry process representative of the local natural language, including abbreviations and jargon.
- Reference Terminology: this is the format used to store data, knowledge information should be stored in this layer
- Aggregate Terminology: different data formats outputs for user analysis

Smith and Ceusters have pointed a capital point for GP/FM; time[15]. Existing biomedical information systems including SNOMED-CT are not concerned by keeping track how entities (such as symptoms or diseases) develop over time. This is the corner stone of the International classification of Primary Care, the tool created by the Wonca International Classification Committee (WICC)[14] for epidemiological purpose since the 70'. ICPC proposes a retrieval methodology based on Episode which means time.

Mapping

Mapping, translation of clinical data into alternative representations without corrupting the primary clinical mission is already done for ICD-9-CM. The task has been completed by mapping ICD9 for billing purpose in 200916. IHTSDO has several teams involved in this issue in all the main fields of classifications like ICD10, NANDA or ICPC. Special Interest Group (SIG) has been launched to study the problem and propose ad hoc solution. The IHTSDO Special Interest Group in Primary Care (SIG PC) has been launched in Copenhagen after an agreement between the World Organisation of Family doctors (WONCA) and IHTSDO. Some Wonca International classification committee members will lead this group. The Family Medicine Research Centre at the University of Sydney, of which leading members are also members of WICC is currently mapping ICPC-2 PLUS to SNOMED to form a first cut Australian GP subset for Primary Care[17]

Recently 2.500.000 diagnostic data collected in electronic patient records throughout 2006 within the County Council of Stockholm and coded in ICD10 (Swedish primary care version) have been mapped to SNOMED CT[18]. The use of Is_a relationships in SNOMED CT aggregated the diagnostic data to 2861 concepts, showing a new, multidimensional view of different specific medical aspects, where every view can be further explored. The authors conclude SNOMED CT provides a different view of diagnoses and health problems on a chapter level, and adds significant new views of the clinical data with aggregations generated from SNOMED CT

Availability of translations

Following a source from Infoway/inforoute Canada, SNOMED_CT translation in French is not yet finished. Inforoute Santé Canada has begun the translation of a list of actives concepts, following determined priorities and actually, only about 10% (30.000 concepts are available in French. From a source in The Netherlands nothing has been translated in Dutch till now but this has to be confirmed. SNOMED-CT is available in several other languages of which Spanish.

6. Does SNOMED CT encompass the field of GP/FM?

Quick look through the Cliniclue browser

Examining the four fields of the Figure 5, one could say that SNOMED-CT succeeds extremely well in the biomedical field. It is probably the most powerful nomenclature for use in pathology, general internal medicine and surgery. A rare disease and related concept as Apoliproteine has 90 occurrences in the Cliniclue browser while the color Blue is quoted 645 times, which is explained by the huge use of this term in tissue stain in Pathology . The field of psychodynamics is clearly dominated by the DSM IV concepts and system of classification, which could be certainly a problem as this particular view of psychiatry and psychodynamic suffers of a story of 'parti pris' and competing interests[19, 20]. Nevertheless, from the Primary Care point of view one will appreciate the

[14] Wonca International Classification Committee (WICC) http://globalfamilydoctor/wicc

147 occurrences of the term 'fear of'. From epidemiological point of view, the vocabulary looks not so developed with 23 occurrences of the concept Public health, 4 for nosocomial and 3 for normal distribution. The anthropological field looks very poor and marked more by common popular view than scientific analysis, particularly when addressing the issue of separating the human beings groups. As usual in medical North American literature the term race is still in use for this purpose despite scientific recommendations21 while the concept of ethnicity suffers from a lack of knowledge. It is quite amazing to read that Belgians is an ethnic group, wedged alphabetically between the Basque and the Black ethnic group (Cliniclue browser). A concept like second generation, which is the corner stone of a consultation in European cities where ethnic groups are so prevalent, is not present in SNOMED-CT terminology behalf for second generation drugs.

Ethnicity : 7 entries
Ethnic group ; 1 entry, defined by *is_a group*
Race : 16 entries
the term Race alone is defined : *is_a demographic history detail*
Race: white : 1 entry : defined : *is_a racial group*
There is no Race : African or black. But well a Black ethnic group.
second generation : 7 entries but only related to the concept of second generation drugs. The concept second degeneration of an ethnic group is absent

Figure 20 Separating the human beings groups in SNOMED-CT

7. Conclusion

While one can acknowledge the fantastic work of aggregating so huge quality products and the high level of expertise involved in it, on can note that, despite to be owned by a non profit organization, SNOMED-CT is a proprietary product which searches an expansion of his market. This implies some rules of confidentiality which are incompatible with Knowledge as a resource of mankind for thos who don't have full access granted. Considering the availability of high level Ontologies in open source[6], it is worth to evaluate also this way of dealing with medical knowledge

SNOMED is still growing. In the 90' Dr C. Jeanty , a Belgian doctor involved in SNOMED use since 1975 was already astonished by the enormous inflation in the terminology of the SNOMED INTERNATIONAL (1993) passed from 44,487 to 132,641 terms[22]. As stated above SNOMED-CT has now 308,000 active concepts, 791,000 active descriptions, organized into 19 hierarchies with more than 951,000 relationships. The great Umberto Eco[23] has written a demonstrative essay "On the impossibility of drawing a map of the Empire on a scale 1 to 1" . Knowledge is not the Empire but it's doubtful that inflation of terms and relations will help to manage it.

8. References

[1] Thomas R. Gruber. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, 1993. http://tomgruber.org/writing/ontolingua-kaj-1993.htm

[2] Van Dormael M. – Médecine générale et modernité. Regards croisés sur l'Occident et le Tiers Monde [Thesis]. Bruxelles: Université Libre de Bruxelles,1995. 286 p.

[3] Greenhalgh T & Hurwitz B. (eds) Narrative Based Medicine (Paperback). BMJ. 1998

---

[4] WangAY, BarrettJW, Bentley T, et al. Mapping betweenSNOMED RTand clinical terms version 3: a key component of the SNOMEDCT development process. Proc AMIA Symp. 2001;741–5. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243316/?tool=pubmed

[5] Joubert M, Abdoune H, Merabti T, Darmoni S, Fieschi M. Assisting the translation of SNOMED CT into French using UMLS and four representative French-language terminologies. AMIA Annu Symp Proc. 2009 Nov 14;2009:291-5. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815379/

[6] Smith B, Ashburner M, Rosse C, Bard C, Bug W, Ceusters W, Goldberg L J, Eilbeck K, Ireland A, Mungall C J, The OBI Consortium, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone S-A, Scheuermann R H, Shah N, Whetzel P L and Lewis S (2007). "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration", *Nature Biotechnology* 25, 1251 - 1255. http://www.obofoundry.org/ontologies.shtml

[7] *Canguilhem G. The Normal and the Pathological*, Cohen (New York: Zone Books, 1991).

[8] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C.Relations in biomedical ontologies. Genome Biol. 2005;6(5):R46. Epub 2005 Apr 28.http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1175958/?tool=pubmed

[9] Robert H. Dolin, Kent A. Spackman & Markwell, D. Selective retrieval of pre- and post-coordinated SNOMED concepts. AMIA *Symp.* **2002**, 210–214. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2244193/

[10] *Wonca International Classification Committee.* Icpc-2-R: International Classification of Primary Care *by 2005.*

11 Rogers J, Bodenreider O. SNOMED CT: Browsing the browsers. In: Proc KR-MED 2008 Representing and sharing knowledge using SNOMED (Phoenix, Arizona) http://mor.nlm.nih.gov:8000/pubs/pdf/2008-krmed-jr.pdf

[12] Cornet R, de Keizer N. Forty years of SNOMED: a literature review. Med Inform Decis Mak. 2008; 8(Suppl 1): S2. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2582789/

[13] Lussier Y. A. and Bourque M.. Comparing SNOMED and ICPC retrieval accuracies using relational database models. Proc AMIA Annu Fall Symp. 1997: 514–518. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233545/

[14] Gambarte ML, Osornio AL, Martinez M, Reynoso G, Luna D, de Quiros FG. A practical approach to advanced terminology services in health information systems. Stud Health Technol Inform. 2007;129(Pt 1):621-5. http://www.ncbi.nlm.nih.gov/pubmed/17911791

[15] Smith B, Ceusters W. An Ontology-Based Methodology for the Migration of Biomedical Terminologies to Electronic Health Records. AMIA Annu Symp Proc. 2005; 2005: 704–708. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560617/?tool=pubmed

[16] HDM Breaking News, November 18, 2009**.** SNOMED CT, ICD-9-CM Map Released. http://www.healthdatamanagement.com/news/-39384-1.html

[17] Wang Y, Patrick J, Miller G, O'Hallaran J 2008. A computational linguistics motivated mapping of ICPC-2 PLUS to SNOMED CT. BMC Med Inform Decis Mak. 2008 Oct 27;8 Suppl 1:S5. http://www.ncbi.nlm.nih.gov/pubmed/19007442?dopt=citation

[18] Vikström A, Nyström M, Ahlfeldt H, Strender LE, Nilsson GH. Views of diagnosis distribution in primary care in 2.5 million encounters in Stockholm: a comparison between ICD-10 and SNOMED CT. Inform Prim Care. 2010;18(1):17-29.

[19] Cosgrove L., Krimsky S., Vijayaraghavan M., Schneider L. Financial Ties between DSM-IV Panel Members and the Pharmaceutical Industry. Psychother Psychosom 2006;75:154–160 http://www.tufts.edu/~skrimsky/PDF/DSM%20COI.PDF

[20] Lane C. Shyness: How Normal Behavior Became a Sickness. Yale University Press; 2007

[21] Witzig R. The Medicalization of Race: Scientific Legitimization of a Flawed Social Construct. Annals of internal medicine. *October 15, 1996 vol. 125 no. 8 675-679. http://www.annals.org/content/125/8/675.abstract*

[22] Jeanty C. Critics of SNOMED INTERNATIONAL [Critique du SNOMED International] in French. Not dated. 1994? http://cybertim.timone.univ-mrs.fr/Members/gbotti/dossier-public/ARTSNOMEDINT/download.

[23] How to Travel With a Salmon & Other Essays by Umberto Eco and William Weaver. Harcourt Brace. 1994. p95

%%%%%%%%%%%%%%%%%%%%%%%%%%%%